The logic of intentional explanation

Kuipers, Theo A.F.

Published in:
EPRINTS-BOOK-TITLE

Publication date:
1985

Communication & Cognition
Vol. 18, No 1/2 (1985), pp. 177–198

THE LOGIC OF INTENTIONAL EXPLANATION*

Theo A.F. Kuipers
University of Groningen

There are two well-known theories of intentional explanation, i.e. explanation of human actions in terms of intentions or goals: the nomological and the semantic theory, with the main proponents Hempel[1] and von Wright,[2] respectively. Both theories are based on the general assumption that an intentional explanation is an argument and the specific assumption that the conclusion of this argument states that somebody performed a certain action. This specific assumption will be called the standard assumption and theories based on it standard theories.

The *nomological* theory pretends to specify the additional conditions an argument should satisfy in order to be acceptable as an intentional explanation. This normative pretension therefore starts from the standard assumption. For the nomological theory the question whether the analysis provides an adequate explication of everyday intentional explanations is of secondary importance, i.e. the explicative pretension does not play a decisive role.

The *semantic* theory is also a standard theory. Although it certainly has normative pretension the explicative pretension at least plays an equal role. One might say that the semantic theory arose as a critical reaction on the explicative value of the nomological theory. However, the semantic theory also does not dispute the standard assumption.

This paper explores the idea that the standard assumption, although it may be a legitimate point of departure for an interesting kind of explanation, is not acceptable as part of an explication of everyday intentional explanations. If this is so then there is, besides the best standard theory, room for a non-standard theory in which the explicative pretension plays the dominant role.

In Section 1 we will evaluate the explicative value of the two standard theories in some detail, leading to a number of 'explicative objections' to both. In Section 2 we will develop our 'non-standard theory': the underlying argument of an everyday intentional explanation is reconstructed as an application of the logical rule of inference called existential generalization with the conclusion that

the action was intentional. It will be shown that such an intentional explanation does not have any of the explicative objections and that it is essentially weaker than a nomological explanation. In Section 3 some attention will be given to additional questions that can be raised when a true intentional explanation has been given in answer to the question of why somebody performed a certain action. Among others, it will be argued that an intentional explanation of an action leaves room for an additional utilistic explanation of the choice of that particular action out of alternatives serving the same purpose.

Finally, in Section 4 we will pay attention to the surprising aspect of our theory that essentially similar explications can be given of other types of explanation about which there is much dispute of whether they are reducible to nomological explanations: viz. functional, teleological and psychological explanations. Instead of being reducible, all these types of explanation turn out to capture the interesting part of otherwise pretentious nomological explanations.

1. The explicative value of the standard theories

In the following evaluation of the two standard theories it is impossible to do justice to all the proposed refinements, but neither is this necessary for our purposes.

The nomological theory construes intentional explanations according to the model of explanation of individual events in the natural sciences. The event to be explained is that somebody x has performed a certain action a, abbreviated as $P(x,a)$, and this should be the conclusion of the argument to be constructed. Three premises are introduced,[3] two of them being specific conditions, the third being a nomological premise. One specific condition, here called the desiderative premise, states that x desired a certain state of affairs or goal g, $D(x,g)$. The other, called the epistemic premise, states that x believed that a was necessary to approach g, $BN(x,a,g)$. Finally, the nomological premise is some kind of universal generalization *(UG)* of the conditional statement 'if $D(x,g)$ and $BN(x,a,g)$ then $P(x,a)$'. It should be a generalization over 'occasions' and it may or may not be a generalization over persons, actions or goals, or over a combination of these. (Note that all specific statements refer implicitly to a certain occasion.)

The result is the following :

Nomological Argument

$D(x,g)$

$BN(x,a,g)$

$\underline{UG \text{ (if } D(x,g) \text{ and } BN(x,a,g) \text{ then } P(x,a)}$)

P(x,a)

It is easy to see that the nomological argument is a valid argument. It is considered to be an adequate or true intentional explanation if, in addition, the premises may be supposed to be true. It is assumed that the specific premises are independent of the (specific) conclusion and hence that the verification of these premises is also independent of the verification of the conclusion[4].

Although the formulations of the specific statements are in the past tense, it would have been possible to present them in the present tense, or even in the future tense. As a consequence, the nomological argument is also a predictive argument; in the past tense it is, more specifically, a retrodictive argument. For this reason the nomological theory is said to be symmetric with respect to explanation and prediction.

In order to assess the explicative value of the nomological theory for everyday intentional explanations we need a first characterization of the latter. An everyday intentional explanation of an action can be represented, in first approximation, in a question- and answer-scheme (QA-scheme) with a standard[5] pattern. The standard question is a 'why-question': 'Why did he/she/you/I do this ?' or, in a general form: 'Why did *x* perform action *a* ?'. Irrespective of the way in which it is obtained, the standard answer to a why-question is of the form '*x* had that intention' or '*x* aimed at that goal' or, more complete, '*x* performed *a* with the intention of approaching goal *g*'.[6]

The answer to a why-question may also be of a different form, for instance, it may state that the behavior of *x* is not adequately described as (an instance of) action *a*. In such a case, however, we would not call it an intentional explanation. In sum, the standard pattern is given by .

QA-scheme (everyday intentional explanation)

Why did *x* perform action *a* ?

x performed *a* with the intention of approaching goal *g*.

The question of the explicative value of the nomological theory can now be put more precisely as: to what extent is it possible to consider the nomological argument as the argument underlying the QA-scheme, if any ?

Let us first ask whether we can retrace the elements of the nomological argument, irrespective of their status as premise or conclusion. The desirative premise is obviously implied by the why-answer. The conclusion of the nomological argument is certainly a presupposition of the why-question and it is explicitly contained, and hence reaffirmed, in the why-answer. The epistemic premise is not explicitly stated in the QA-scheme, but it seems defensible that at least a weakened form (e.g. '*x* believed that *a* was useful to obtain *g*') is a meaning component of the why-answer.

However, the fourth element of the nomological argument, the nomological premise, does not seem to be retraceable in the QA-scheme at all. Formulated as an explicative objection to the nomological theory we can put it in

the following way: what does it matter what somebody (or other people) does (do) on similar occasions, the question is why somebody did what he did on this particular occasion. We will call this objection the *irrelevance*-objection to the nomological premise and hence to the nomological theory. It cannot be stressed enough that this objection is an objection to this theory only if it is (also) considered to be an explication of everyday intentional explanation.

The second explicative objection is the *symmetry*-objection. According to the nomological theory there is no essential difference between intentional explanation and 'intentional prediction' (including retrodiction). This objection may seem odd at first sight. If prediction is an important aspect of the sciences and if common sense is at least to some extent an anticipation of scientific knowledge, how is it then possible to call the transformability of intentional explanations into predictions an objection ? On second thoughts, however, this is not convincing. Many historians for instance would not claim that they aim at retrodictions of human action, some of them may even be proud of that. Similarly, biologists do not aim with their functional explanations of organs in organisms at a retrodiction of the presence of these organs. Of course, this does not exclude the important role that predictions play in testing their explanations. Analogously, we seldom claim in everyday life that an intentional explanation could have predicted that the person in question would perform the particular action, notwithstanding the fact that an (hypothetical) explanation implies predictions (although perhaps only 'weak' predictions).

The third explicative objection is strongly related to the foregoing one. It concerns the standard assumption that the conclusion of the underlying argument is that somebody performed a certain action. If there is an underlying argument at all, its conclusion must be different from the standard assumption. That somebody performed a certain action is a presupposition underlying the whole QA-scheme. Although it is recognizable, it is not at all disputed. Hence, it should enter as a premise. To put in it a different way, intentional explanations are essentially explanations *ex post actu,* (hence, they cannot be transformed into retrodictions). This objection will be called the *conclusion*-objection.

The last explicative objection to be considered is the *necessity*- or *alternatives*-objection. The epistemic premise stated that x believed that a was *necessary* to obtain g. It certainly occurs sometimes that this premise may be considered to be true, i.e. situations in which people think, rightly or wrongly, that they should do *this* to achieve *that*. But it occurs more frequently that only a weaker premise can be justified, namely in the case that the actor believes that there are alternative actions serving the same purpose. Within the nomological theory a first step[7] is to introduce a 'disjunctive action' $d(a_1, a_2, .,a_n)$ to be substituted in the epistemic premise as well as in the nomological premise and

the conclusion. It is clear that this procedure weakens the explanatory (and predictive) power of the nomological argument with respect to a particular action a_i. It explains, so to speak, the particular action 'only up to alternatives'.

The objection is of course the following. In everyday intentional explanation alternative actions do not play any role, even if there are obvious alternative actions known to the actor. To the question 'Why did he open the window ?' the answer 'He wanted to cool the room' provides a perfect explanation. Only if we are interested in the alternative action of, say, lowering the heating might we ask the additional question 'Why did he not lower the heating ?'. But then we have a well-distinguished new why-question.

If there is a problem of non-uniqueness in everyday intentional explanations it is not the problem of alternative actions but the problem of different goals. That is, there may be different intentional explanations of the same action, leading to the problem of the relation between them.

This completes our evaluation of the explicative value of the nomological theory. In the light of the four objections we conclude in particular that the nomological argument is not the underlying argument, if there is one, of everyday intentional explanations.

We will now turn to the evaluation of the semantic theory. As has already been said this theory may be considered as a reaction to the explicative shortcomings of the nomological theory. It will turn out that the irrelevance- and the symmetry-objection do not apply to the semantic theory. However, the conclusion-objection remains (it is a standard theory), as well as the alternatives-objection. Moreover, there is a new objection, called the magic-objection.

The semantic theory is usually presented in the form of the so-called practical syllogism, i.e. the nomological argument without the nomological premise:

Practical Syllogism
D(x,g)
BN(x,a,g)
P(x,a)

It is easily seen that the practical syllogism is, in the form presented, not a logically valid argument. But an argument may be analytically valid[8] : valid on the basis of meaning relations between the terms used in the premises and the conclusion. An analytically valid argument can be transformed into a logically valid argument by adding appropriate semantic or meaning postulates specifying the meaning relations. The semantic theory states that the practical syllogism is such an analytically valid argument. The corresponding meaning postulate is, qua form, the same as the most farreaching nomological premise:

MP For all persons x, all actions a, all goals g and all 'occasions':
if $D(x,g)$ and $B(x,a,g)$ then $P(x,a)$.

The status of MP, however, is completely different from that of a nomological premise. The truth of the latter is considered to be a matter of facts, not of meanings. On the other hand the ttruth of MP is considered to be of the opposite nature: not a matter of facts but of meaning.

Let us first of all note that the irrelevance-objection does not apply to MP even though MP is not explicitly recognizable in the QA-scheme. The reason is that meaning relations are to be expected in everyday arguments without some explicit indication. It is also clear that the practical syllogism transforms with MP into a logically valid argument.

However, we will call MP *magical* for the following reason: it presupposes that, what is in our opinion a false claim, natural language is such that particular desires and beliefs imply, as a matter of meaning, a particular action. To be sure, the problem is not that some primarily mental concepts have behavioral connotations, e.g. being hot-tempered, but that statements using only such concepts, even if they are quite specific, might imply a *specific* action.

The magic-objection becomes even more clear if we see how MP succeeds in avoiding the symmetry-objection. The predictive power of the nomological argument arose from the independent verifiability of the premises. According to the semantic theory however, sufficient verification of the premises *D(x,g)* and *BN(x,a,g)* is only possible, in view of the nature of a meaning postulate, by verification of the conclusion *P(x,a)*. In other words we are only justified in speaking of genuine 'desiring' and 'believing to be necessary' if the relevant action is performed. Hence, the practical syllogism cannot be used as a predictive argument.

It may seem surprising that proponents of the semantic theory like to stress this asymmetric character of their theory, appealing to the verification circularity, but are at the same time reluctant to draw the conclusion that intentional explanations are essentially *ex post actu*.

This reluctance may become understandable if we realize that they stick to the standard assumption that *P(x,a)* is the conclusion of the underlying argument. As long as this assumption is made an explanation *ex post actu* will readily have an *ad hoc* air. However this may be, the conclusion-objection applies to the semantic theory for the same reasons as it applied to the nomological theory.

It is also not difficult to check that the same holds for the alternatives-objection. This completes the evaluation of the explicative value of the semantic theory, and we conclude that the practical syllogism is also not the argument underlying everyday intentional explanations.

If we put aside for a moment the explicative aim the natural question is which of the two theories presents the best account for a concept of explanation based on the standard assumption. In this respect the nomological theory seems

to be the best candidate. The conclusion-objection no longer applies to either theory. The irrelevance-objection and the symmetry-objection do not now apply to the nomological theory. The alternatives-objection would, if it remains an objection at all, equally apply to both theories. But the magic-objection remains to be applicable only to the semantic theory in the following form: in concept formation we should avoid magical meaning relations of the kind of the meaning postulate assumed in the semantic theory.

2. Explication of everyday intentional explanation

In the preceding section we have already formulated the standard question and answer pattern for everyday intentional explanation. It read:

QA-scheme

Why did x perform action a ?

x performed a with the intention of approaching goal g.

It is easy to assess that we use this type of explanation frequently and easily in everyday life. Moreover, from the QA-scheme it is immediately clear that an intentional explanation will be called true if and only if the why-answer is true. There would be not much more to explicate about intentional explanations if there were not associated with it some intuitions which are not directly clear from the QA-scheme. That is, the primary task of explication is the explication of these associated intuitions.

We will consider especially two, more or less competing, intuitions: the argument-intuition and the understanding-intuition.

The *argument-intuition* we have already met as a general assumption of the two standard theories. It states that there is a valid argument underlying intentional explanations, such that the explanation is true if and only if the premises of the argument are true. We do not claim that everybody will share, or needs to share, this intuition (see below). But it is undoubtedly shared by many philosophers. On the basis of the preceding section we may already lay down some negative (related) characteristics of this argument:

1) its main conclusion should not be that somebody performed a certain action (this should be a kind of presupposition of the whole argument so that the explanation becomes *ex post actu* in an essential sense),
2) it should not be transformable into a prediction or retrodiction (i.e. it should be asymmetric with respect to explanation and prediction),
3) it should not have a (crucial) nomological premise, to avoid irrelevance,
4) meaning postulates, if necessary to obtain logical validity, should not be magical,
5) alternative actions serving the same purpose should not play a role (on the other hand, the phenomenon of different (true) explanations of the

same action should be possible).

The *Verstehen- or understanding-intuition* may be presented as a criticism of the argument-intuition. The intuition of an underlying argument, is, according to this criticism, just a prejudice, probably originating from the (false) analogy with explanation of individual events in the natural sciences. What an intentional explanation does is to make the action understandable to the questioner: although this does not seem to be an argument, this is the heart of the matter, according to the understanding-intuition. Of course, this possible motivation for the understanding-intuition does not exclude an explication of both intuitions in a harmonious way.

We will start with the explication of the argument-intuition. Let us analyse in more detail the meaning of a why-answer, or, more generally, of a statement of the form

PI(x,a,g) x performed a with the intention of approaching g

where 'x' represents 'person x', 'a', 'action a' and 'g', 'goal g'.

It seems plausible to distinguish at least the following three meaning components:

D(x,g) x desired g
BU(x,a,g) x believed a to be useful to approach g
P(x,a) x performed a.

We have already met these three components in the standard theories, albeit the epistemic component *BU(x,a,g)* is now weaker than the epistemic premise *BN(x,a,g)*. However, this weakening is unavoidable for *PI(x,a,g)* evidently does not imply that x believed a to be *necessary*. Neither does it imply that x believed a to be sufficient. Of course, we do not want to say that there are no occasions on which somebody believes that a certain action is necessary or sufficient for a certain goal, but only that neither of them belongs to the meaning of *PI(x,a,g)*.

In terms of probabilities it is plausible to interpret *BU(x,a,g)* simply as the fact that, for x, the subjective unconditional probability that g will occur is lower than the subjective probability of this event under the condition that x performs action a. It is easy to check that, under very general conditions, this is equivalent to the fact that the subjective probability that g will occur if x does not perform a is lower than if x performs a.

Suppose now for a moment that x believed that a was sufficient to approach g, i.e. the subjective probability that g will occur if x performs a is 1. A special kind of this will occur if g is explicitly mentioned in the description of a, e.g. 'to open the door' and 'having the door opened'. In this case we say that *BU(x,a,g)* and, hence, *PI(x,a,g)* are trivial. *In what follows we will always presuppose the non-triviality of these statements.*

In our opinion the three components do not only belong to the meaning

of *PI(x,a,g)*, they are, with some qualification, also exhaustive, i.e. their conjunction is equivalent in meaning to *PI(x,a,g)*. The qualification is of temporal nature. Suppose *D(x,g)* or *BU(x,a,g)* became true only after *P(x,a)*, i.e. *x* got the desire, or the belief, only after actual performance of the action. In this case we would be reluctant to call *PI(x,a,g)* true. Hence, without mentioning it further, we will assume that the desire and the belief precede the action in a true statement of the form *PI(x,a,g)*.

In terms of the conjunction

(1) $D(x,g) \,\&\, BU(x,a,g) \,\&\, P(x,a)$

we may summarize the foregoing in the following meaning postulate:

MP.1 *PI(x,a,g)* if and only if (1)

In the line of the nomological theory we assume also that the 'mental' components *D(x,g)* and *BU(x,a,g)* are independent of the 'action' component *P(x,a)*. In this way we renounce of course a variant of the crucial meaning postulate of the semantic theory, i.e. excluding *D(x,g)* and *BU(x,a,g)* to be true and *P(x,a)* false. But, to be honest, due to the weakening of the epistemic component this variant lacks any plausibility. Another combination of truth-values which might seem problematic at first sight is *P(x,a)* and *BU(x,a,g)* true and *D(x,g)* false. Note, however, that it makes perfect sense to talk about undesired (or even unintended) but expected consequences of an action.

An important aspect of *PI(x,a,g)* and its components is that they are all (singular or) *specific* statements. All of them refer to a specific person and, as far as relevant, to a specific action or goal.

Now the question arises whether *PI(x,a,g)* has also interesting *unspecific* meaning components, i.e. statements that follow from it perhaps, with the aid of some new meaning postulates, and that generalize in some way or other over persons, actions, or goals. Universal generalizations are of course out of the question. On the other hand a (large) number of existential generalizations are allowed. In the context of intentional explanation, where the why-question presupposes the truth of *P(x,a)*, one existential generalization of the additional content of *PI(x,a,g)*, i.e.

(2) $D(x,g) \,\&\, BU(x,a,g)$

will turn out to be of crucial importance, namely[9] (with y as variable for goals),

(3) $(Ey)\,(D(x,y) \,\&\, BU(x,a,y))$

Let us write down the conjunction of (3) and *P(x,a)* :

(4) $P(x,a) \,\&\, (Ey)\,(D(x,y) \,\&\, BU(x,a,y))$

The verbal formulation of (4) is of course

PI(x,a) : x performed *a* intentionally

where 'intentional' is used in the sense of 'with some intention' or 'with the intention to approach some goal'. That is, the following meaning postulate is plausible:

MP.4 *PI(x,a)* if and only if (4).

From (4) and MP.4 it is clear that *PI(x,a)* should be distinguished from '*x* intended to perform *a*' or '*a* was an intended action of *x*'. These expressions may already be taken as analytic consequences of *P(x,a)*, i.e. actions are by definition intended actions.[10] This aspect is however not relevant for us.

Consider now the statement

IR(x,a) : *a* was intentionally relevant for *x*

For this statement the following meaning postulate is plausible:

MP.3 *IR(x,a)* if and only if (3)

Now it is natural to introduce also the somewhat clumsy statement

IR(x,a,g) 'a was intentionally relevant for *x*'s desire *g*

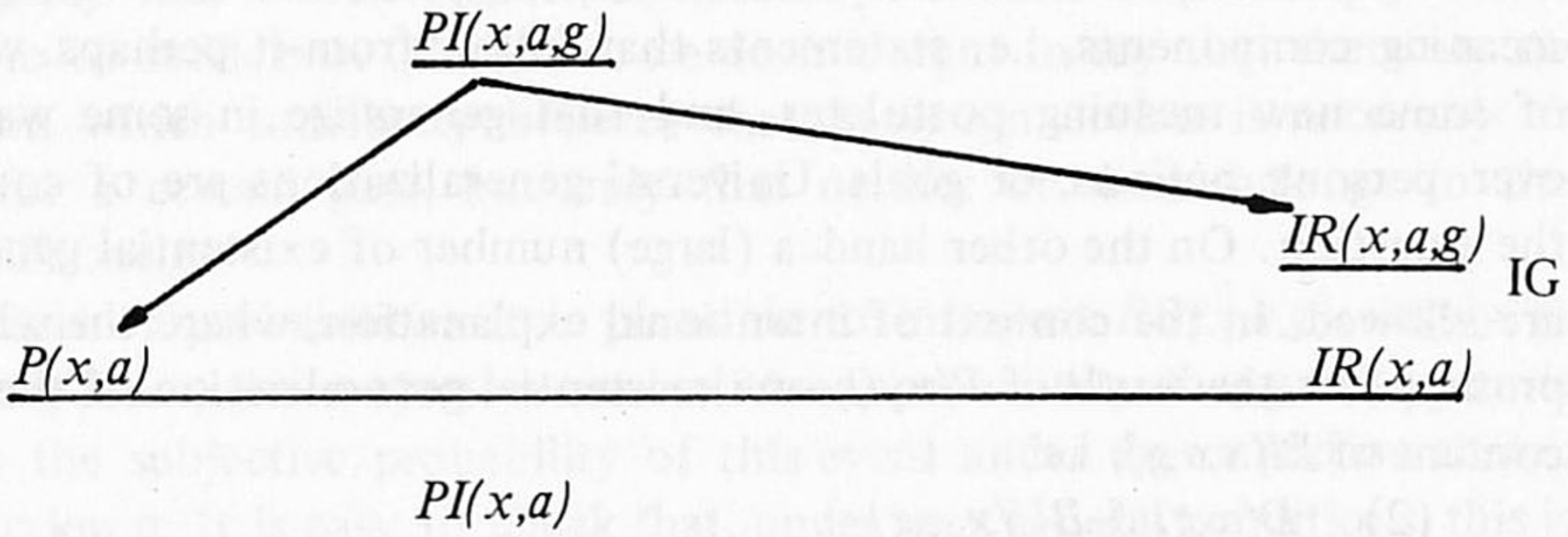
with the corresponding meaning postulate

MP.2 *IR(x,a,g)* if and only if (2)

The transition from (2) to (3) is a straightforward application to (2) of the logical rule of inference called existential generalization, and the result (3) is called an existential generalization of (2). Because of MP.2 and MP.3 we may use the same terminology for the transition from *IR(x,a,g)* to *IR(x,a)*. For obvious reasons we will talk in the present context about intentional generalization (IG).

On the basis of the four meaning postulates and the standard rules for conjunction we can present now the following deductive argument

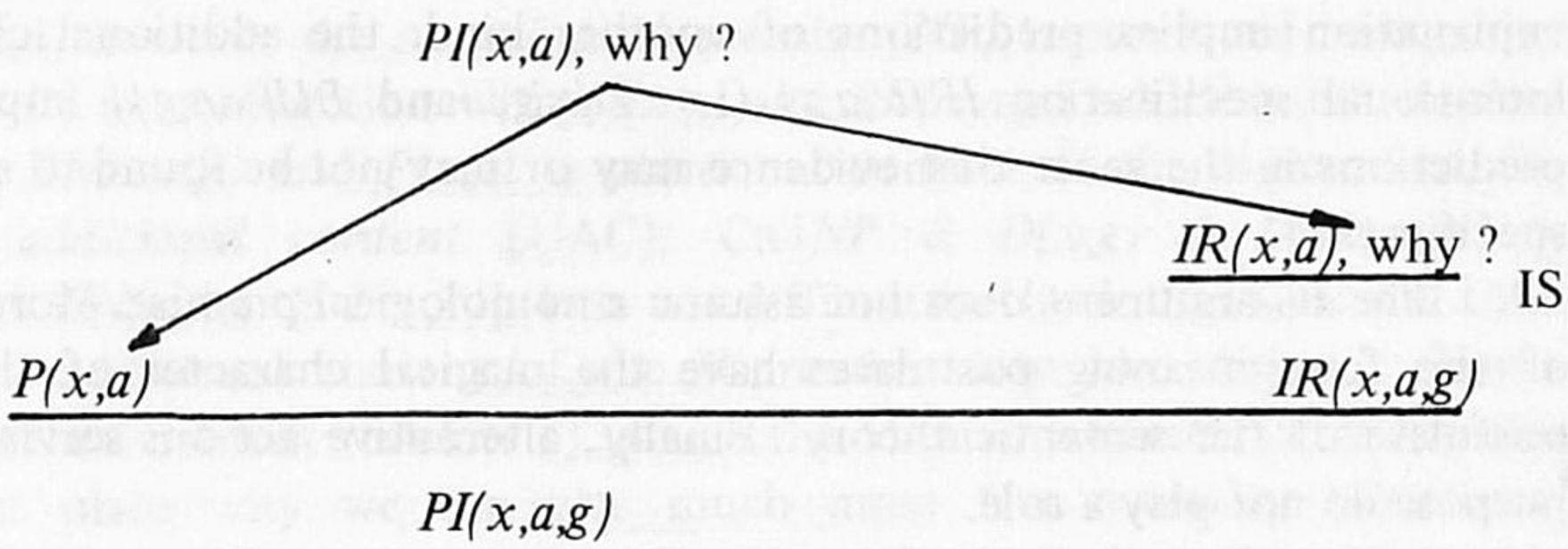
IG-argument



Our central claim is that the argument giving rise to the intuition that there is an argument underlying everyday intentional explanations is this IG-argument. To be precise, our claim is that the QA-scheme can be reconstrued as an IG-argument read upside down.[11]

To argue this let us start from the why-question: 'Why *P(x,a)* ?'. The apparent assumption of this question is not only *P(x,a)* but even *PI(x,a)*, i.e. *x* performed *a* intentionally. What the questioner wants to know is why this is so, i.e. *PI(x,a)*, why ? According to MP.4 we may split *PI(x,a)* into *P(x,a)* and

IR(x,a). It is clear that the questioner does not dispute *P(x,a)*. Hence, the why-question is more particularly directed to *IR(x,a)*, i.e. the questioner wants to know why this is so, i.e. *IR(x,a)*, why ? The why-answer introduces now a transition from *IR(x,a)* to *IR(x,a,g)*, which we will call an *intentional specification* (IS). Of course, IS is not a logical rule of inference (it should not be confused with the so-called rule of existential instantiation). IS introduces a specific goal about which *IR(x,a,g)* is claimed to be true. Combining the intentional specification again with *P(x,a)* we obtain of course, *PI(x,a,g)*. In total we get the following 'deep-structure' behind a QA-scheme :



It will be clear that the IG-argument arises from reading the IS-scheme upside down or, to put it differently, the IG-argument can be attached to the IS-scheme and it brings us back to the initial assumption of the IS-scheme: *PI(x,a)*.

An obvious objection to our explicative claim is the following: it does not make much sense to attach the IG-argument to the IS-scheme for the conclusion of the former, *PI(x,a)*, is supposed to be the crucial assumption of the latter. The point is of course that this assumption, or more specifically, *IR(x,a)* might be false in which case intentional specification and *hence* intentional generalization would not be possible.

To put it differently, verification of *IR(x,a)* is only possible by intentional specification and subsequent generalization. In this respect the other assumption, *P(x,a)*, is different: intentional specification is certainly not the way to verify it.

Yet another way to put the crucial point is as follows: if we agree that intentional specification is the central *synthetic* (or inductive) aspect of an intentional explanation, and this seems very reasonable, then it is also quite reasonable that intentional generalization is the central *analytic* (or deductive) aspect of an intentional explanation.

We have called an intentional explanation formulated in a QA-scheme

true if and only if the why-answer was true. It is clear that the deep structure, the IS-scheme, is in accordance with this definition. Consequently, this truth definition corresponds to saying that the (conjunctive) premise of the IG-argument is true.

It is easy to check that the IG-argument has all five negative characteristics specified at the beginning. The crucial one is the first: the main conclusion of the IG-argument is not that x performed a. The main part of the conclusion is obviously $IR(x,a)$. If combined with $P(x,a)$ we get the complete conclusion, but $P(x,a)$ only enters as part of the initial premise, i.e. as an assumption. In this sense the explanation is *ex post actu*, i.e. the conclusion qualifies the premise $P(x,a)$. From this it follows immediately that the IG-argument cannot be used as a prediction of $P(x,a)$. Without going into details it is clear that an intentional explanation implies predictions of another kind: the additional claim of the intentional specification $IR(x,a,g)$ (i.e. $D(x,g)$ and $BU(x,a,g)$) implies (weak) predictions in the sense that evidence may or may not be found to support this specification.

The IG-argument does not assume a nomological premise. Moreover, none of the four meaning postulates have the magical character of the meaning postulate of the semantic theory. Finally, alternative actions serving the same purpose do not play a role.

Let us now turn to the understanding-intuition. From the foregoing it is clear that we do not assume that every human action can be explained intentionally, i.e. that there is a true intentional explanation for it. This does of course not imply that no other type of explanation can be given for it, e.g. a 'psychological explanation' in terms of unconscious goals. The expression 'an action is understandable' can now be used in a strict, and in a broad sense. In the strict sense we equate 'the action is understandable' with 'the action is intentional'. In the broad sense the latter is a special case of the former, i.e. 'the action is intentionally understandable'. However this may be, from a true intentional explanation we may conclude that the action was understandable in both senses of the term. And even stronger, the intentional specification provides the grounds for saying that we have actually understood the action. Hence, the understanding-intuition can be explicated in complete harmony with the explication of the argument-intuition.

An important remaining question is the relation between an intentional explanation in our sense and explanations according to the nomological and the semantic theory. We will call explanations of the same action *corresponding* if they refer to the same goal and, hence, use the same desiderative premise.

With respect to the nomological theory there is a straightforward answer. A nomological (N-) explanation is stronger than the corresponding intentional explanation in the following sense: the premises of the nomological argument

imply more than the premises of the corresponding IG-argument (and hence they imply its conclusion). First we show that the N-premises imply the three IG-premises:

- *D(x,g)* is an explicit premise of the N-argument,
- *BU(x,a,g)* is, without doubt, analytically implied by *BN(x,a,g)*,
- *P(x,a)* is, by definition implied by *D(x,g)*, *BN(x,a,g)* and the nomological premise.

But the N-premises are even essentially stronger, and even in two, quite different, respects. First, *BN(x,a,g)* is obviously much stronger than *BU(x,a,g)*. We define the *specific additional content* (SAC) as follows: Cn[*D(x,g)* & *BN(x,a,g)*] – *Cn*[*D(x,g)* & *BU(x,a,g)*], where *'Cn'* indicates 'class of analytic consequences'.

Second, however the nomological premise *(NP)* is construed, everything implied by it, in conjunction with *D(x,g)* and *BN(x,a,g)*, which is not implied by *D(x,g)*, *BN(x,a,g)* and *P(x,a)* is additional content of the N-premises, the *unspecific additional content* (UAC): Cn[*NP* & *D(x,g)* & *BN(x,a,g)*] – *Cn*[*D(x,g)* & *BN(x,a,g)* & *P(x,a)*]. It is not difficult to check that SAC and UAC partition all additional content of the N-premises into two disjunct classes.

Additional content restricts of course the 'domain of truth'. This explains in the first place why we can give much more true everyday intentional explanations than true nomological explanations. Secondly it explains the origin of the explicative objections against the nomological theory: this theory asks too much.

For the comparison with explanations according to the semantic theory we can now be shorter. Without the magical meaning postulate the premises of the practical syllogism (PS) only imply the IG-premises *D(x,g)* and *BU(x,a,g)*, and with respect to these the additional content coincides with SAC of the N-premises. The PS-premises imply, in addition, *P(x,a)* only if we accept the magical meaning postulate. Hence, *without* this postulate, PS is an *in*valid argument, incomparable to the IG-argument. *With* this postulate it is valid and stronger than the IG-argument: stronger, in two respects: SAC and the magical meaning postulate.

We conclude this section with a terminological remark. Intentional explanations are also called *rational* explanations. This terminology is however highly misleading for it is likely to confuse two things: reasons and good reasons. The term 'reason-giving-explanations' does not have this ambiguity. The crucial point is that it should be possible to call an action intentional but not (very) rational. For example, for an atheist the praying of a believer may be intentional, but not rational. The general point is that beliefs may or may not be rational. Moreover, even desires may or may not be rational.

3. Additional why-questions

A true intentional explanation of an action does not, of course, imply that there are no questions left. On the contrary, a true intentional explanation is the natural point of departure for new why-questions. To get natural formulations we will rephrase *PI(x,a,g)* as : '*x* aimed, with *a*, at *g*' and introduce the plausible existential generalization for it, i.e., '*x* aimed at goal *g*', with *z* as variable for actions :

A(x,g) if and only if *(Ez) PI(x,z,g)*

The first two questions now are

Q1 Why did *x* aim at *g* ?

Q2 Did *x* aim only at *g* ?

The first question leads immediately to a natural extension of intentional explanations of actions to intentional explanations of goals. In informal terms we get '*x* aimed at *g*' because '*x* desired *g*' and '*x* believed that *g* was useful for *g**', i.e. *AI(x,g,g*)* : '*x* aimed at *g* with the intention of approaching *g**'. It is clear that the same type of IS-scheme and hence IG-argument can be formulated as before.

Suppose now that *PI(x,a,g)* and *AI(x,g,g*)* are true. Does it follow that *PI(x,a,g*)* is also true ? It is easy to check that this new explanation of *P(x,a)* is true if and only if, in addition to *PI(x,a,g)* and *AI(x,g,g*)*, it is also true that *BU(x,a,g*)*.

Consider, however, the following principle:

PTU if *BU(x,a,g)* and *BU(x,g,g*)* then *BU(x,a,g*)*

called the principle of transitivity of usefulness. Note that PTU can easily be extended to cases where '*a*' is replaced by a third goal. Now PTU will not be a universally true principle; it comes close to the principle that people always believe the logical consequences of their beliefs, which is certainly false. Hence, *PI(x,a,g*)* need not be true if *PI(x,a,g)* and *AI(x,g,g*)* are true.

We will say that two intentional explanations of an action, e.g. *PI(x,a,g)* and *PI(x,a,g*)*, are *directly related* if one of the goals can be explained by the other, e.g. *AI(x,g,g*)*.

If two intentional explanations of an action are not directly related they need not be totally unrelated: they may be indirectly related by a third one. That is, $PI(x,a,g_1)$, $PI(x,a,g_2)$ and $PI(x,a,g_3)$ such that $AI(x,g_1,g_3)$ and $AI(x,g_2,g_3)$.

Answers to the second question ('Did *x* aim only at *g* ?') are now easy to give.[12] It may very well be that *x* aimed at more than one goal, and this will usually be the case. However, most of them will be related. The idea that there is *unique* intentional explanation, in many cases, can now be explicated: all goals are related in a special way.

Consider all g for which $PI(x,a,g)$ and let AI constitute a partial ordering of these goals, 'starting' with one or more goals which are even linearly ordered by AI, and so that for any of them, say g, and for any one of the remaining goals, say g^*, $AI(x,g,g^*)$ holds true. If there is such a unique 'goal-chain' at the beginning we can talk about *the* intentional explanation of the action (and of the goals in the chain, except the last one), meaning by it the specification of the first (the next) goal in the chain. But in a more liberal way of speech we will accept the specification of any member of such a chain as *the* true intentional explanation.

Returning to the first question we hasten to add that intentional explanations of goals will bring us, after some steps, to goals for which the interesting question is no longer 'Why did x aim at that goal ?', but 'How did it come about that x wanted that goal ?'. Then we get a transition from intentional explanations of goals to 'historical explanations'[13] of goals.

More or less the same transition arises in the context of the third question raised by an intentional explanation:

Q3 Why did x believe a to be useful for g ?

Analogous to intentional (or goal-) explanations of actions and goals we can construe 'belief-explanations' of beliefs of the form $BU(x,a,g)$, viz. x believes, say, B and x believes that B supports (the belief) that a is useful for g. It is interesting to note that the relation of support is, in probabilistic presentation, opposite to the relation of usefulness. Denoting the belief that a is useful for g by b it is plausible to assume that 'x believes that B supports b' implies $p_x(b) < p_x(b/B)$, whereas $BU(x,a,g)$ is probabilistically interpreted as $p_x(g) < p_x(g/a)$,, i.e. the role of B and b corresponds to that of a and g respectively, whereas we might expect the reverse correspondences.

Again we have the situation that belief-explanations of beliefs will bring us to beliefs for which the interesting question is no longer why x believed it, but 'How did x come to believe it ?', i.e. the quest for a historical explanation.

The last specific question .we will consider, to which a true intentional explanation gives rise, is very crucial for our explication:

Q4 Why did x choose a among the alternatives ?

In Section 1 we saw that the two standard theories were both subject to the alternatives-objection. Our claim was that alternatives do not play a role in an intentional explanation. Our explication obviously has this negative characteristic, but it remains to be shown how alternatives should be treated as soon as they come into the picture, as is the case in Q4.

Let us start with the intentional explanation $PI(x,a_1,g)$ of $P(x,a_1)$, implying $D(x,g)$ and $BU(x,a_1,g)$. Suppose now that also for the (possible) actions $a_2, \ldots,a_n$ holds that $BU(x,a_i,g)$. Let a_o denote 'doing nothing' and let $BU(x,a_o,g)$ be false. Note that this is a non-trivial assumption.

From these assumptions it follows that we would be able to explain all 'positive' actions in terms of the same goal *g*. But *x* did choose a_1, why ?

In order to answer this question we will postulate quantitative utilities *(u)* and probabilities *(p)* for convenience, but only use comparative relations between them.

We start with some obvious interpretations, leaving out the reference to *x* in quantitative terms:

$D(x,g)$: $u(g) > u(\text{not-}g) =_{df} O$

$BU(x,a_i,g)$: $p(g) < p(g/a_i)$ $\quad i=1,2,\ldots,n$

$\text{not-}BU(x,a_o,g)$: $p(g) \geqslant p(g/a_o)$

In addition to these assumptions it is reasonable to assume that each of the positive actions costs more or less effort, i.e. they have, compared with a_o, negative utility

$u(a_i) < u(a_o) =_{df} O$ $\quad i=1,2,\ldots,n$

One standard decision theoretic question now is whether or not the expected utility of a_i, i.e.

$$\begin{aligned} EU(a_i) &= u(g)p(g/a_i) + u(\text{not-}g)p(\text{not-}g/a_i) + u(a_i) \\ &= u(g)p(g/a_i) + u(a_i) \end{aligned}$$

was maximal for a_1. Note that if this is so, it implies also that $EU(a_o)$ $(= u(g)p(g/a_o))$ was smaller than $EU(a_1)$.

Suppose now that *x* made up his mind before performing a_1, considered the alternatives $A = (a_o, a_1, \ldots, a_n)$, attached utilities and probabilities as suggested, and calculated the expected utilities of each action and reached the conclusion that $EU(a_1)$ was maximal and decided to aim, in choosing, at maximization of expected utility and actually chose a_1 and performed it. Now it is natural to say that the intentional explanation of $P(x,a_1)$ is still in terms of *g*, i.e. $PI(x,a_1,g)$, and that *x*'s choice of a_1 out of *A* was also intentional, but now not directed to *g* but to maximizing expected utility. In other words, in this way we arrive at a natural distinction between an intentional explanation of an action and an intentional explanation of 'choosing one out of the alternative actions', i.e. of choice. Intentional explanation of choice will be called utilistic as soon as the actor uses some kind of utility decision rule.

In our opinion the success of decision theory, which is of course mainly interested in utilistic explanations of (and devices for) choice, has unfortunately led to the theoretical neglect of intentional explanation of actions. Standard presentations of decision theory do not usually start from a particular goal, but from incompatible possible outcomes, which obscures the fact that in many cases there is really one goal at stake, that will or will not be achieved. As far as this is true it may very well be that scientific interest in particular actions of human beings, e.g. in history, is more directed to intentional explanation of such actions rather than to utilistic explanations of the choice of them. On the other hand,

in contexts where scientific interest is mainly directed to collective effects of individual actions, e.g. economy, sociology and politicology, it is reasonable to expect that the emphasis will be on utilistic explanations of choices.

These differences in emphasis will partly be due to the fact that there may be a true intentional explanation of an action but not a true intentional, let alone a true utilistic, explanation of the choice of that action. E.g. somebody may jump into the water to rescue someone else, without any deliberation of alternative actions serving the same purpose. Conversely, in a complex decision situation, with many possible actions and incompatible outcomes, it is hardly possible to talk about the intention with which the chosen action is performed, although the choice may be perfectly intentional in a utilistic sense.

We conclude this section with a general question raised by true intentional explanations. If, as we have suggested, there are so many true intentional explanations in everyday life, why are they not considered as belonging to our stock of scientific knowledge ? As with a lot of other (true) information (and intentional explanations just provide a special type of information) its neglect must be because this information is, in general, not considered to be scientifically interesting. An important exception is the science of history where many intentional explanations are considered interesting. As is clear from our exposition, historians should not bother about the well-known nomological criticism. If they support their hypothetical intentional explanations with evidence they explain actions in a significant and sound way. Attempts to strengthen such explanations to nomological explanations will, as a rule, not lead to interesting laws[14]. This does not of course exclude that the historian is interested in statistical information, e.g. of the form: in that period and region most people doing such and such did that with the intention of so and so. Such information may suggest intentional explanations in particular cases. Note that this use of statistical information differs from the imperfect form of nomological explanation, so-called statistical explanation.

4. Extrapolations and speculations

There is not only much dispute in the philosophy of history but also in the philosophy of other sciences as to whether the types of explanation which occur in these sciences are, or are not, of a different nature than the nomological explanations in the natural (non-life) sciences. In particular one may think of functional and teleological explanations in biology and, in addition, psychological explanations in psychology and psychiatry. Among the scientists who use such explanations, and most of them do, many have a bad conscience about it.

In this final section we will argue for the positive claim that such non-nomological explanations have a sound logical structure. In contrast to most

defences of this claim we do not argue that these explanations compete with corresponding nomological explanations. On the contrary, they are, though highly informative, not so pretentious.

We will indicate only some general lines, leaving particular questions open for further research. The basic idea is of course that nomological explication of these, prima facie, non-nomological explanations will meet the same objections as intentional explanations, and that proper explication will reveal that they all have the same underlying argument as we found in the case of intentional explanation, viz. existential generalization.

Let us start with so-called functional explanation in biological anatomy and physiology. Why do certain species of animals have lungs ? Or, equivalently, what is the function of lungs (if the species has lungs) ? A rough, but essentially true, answer is that the function of lungs is to enable air-exchange between organism and environment. The argument is again existential generalization and it is obtained by reading the explanation upside down :

the (or a) function of lungs is air-exchange

lungs are functional.

Of course, this explanation leads to the question: what is the function of air-exchange ? Via a chain of related functions we will finally arrive at the function of the maintenance of life. Of course, we want such a chain to be as detailed as possible, i.e. we want as specific functions as possible.

It is clear that the most important problem for further research is the analysis of the meaning of specific functional statements, for only such an analysis can make clear how such statements, and hence how such explanations can be tested. To be sure, the articulation of the meaning of specific functional statements does not seem to be as easy as in the case of specific intentional statements. But, whatever the meaning is, it will not affect the (sound) nature of the underlying argument. Moreover, it is a safe risk to conjecture that the premises of a nomological explanation in functional terms (with the standard conclusion that certain animals have lungs) will also imply the specific functional statement. Consequently, a functional explanation will be weaker than a nomological explanation in functional terms.

It is however unlikely that the biologist will be interested in a nomological explanation (hence retrodiction) of the fact that these animals have lungs. He happens to have observed that these animals have certain organ-like parts, called lungs, and in view of the fact that most, or at least many, organ-like parts have been found to have specific functions, he asks for the specific function of lungs. Of course, our biologist will be interested in related questions, such as 'How does the process of air-exchange by the lungs work ?' and 'How did the lungs originate phylogenetically ?'. As to the first question some (causal-) nomological

account can be expected. As to the second question the answer should of course be given in terms of the theory of evolution: functions (e.g. functional organ-like parts, i.e. organs) originate from genetic mutation and natural selection on the basis of survival value. In general, it will not be very interesting to lay down such a historical explanation in detail. However this may be, such a historical explanation seems to be the proper treatment of the question a nomological explanation would try to answer.

It is also clear that aspects of the answers to both questions will enter into an explication of the meaning of specific functional statements, i.e. regularities will be involved relating the operation of lungs to survival-value, but this will not bring us to a nomological retrodiction of the existence of lungs, i.e. the regularities will not (be able to) play the role the nomological premise is supposed to play in a nomological argument.

The above account can of course be extrapolated to all kinds of organs and processes in animals (and plants). In the case of processes, e.g. breathing, we are perhaps more inclined to talk about teleological explanations, but the structure of the underlying argument is the same: this process is directed to that goal, hence, this process is goal-directed.

Also the extrapolation to instinctive (outer-) behavior is plausible. If instinctive behavior happens to seem dysfunctional it is likely that the current environment differs in some relevant aspect from the environment in which the natural selection operated. The same holds true of course for rudimentary organs, etc.

There seems also to be no fundamental difference between functional explanation of instinctive behavior and learned behavior, apart from the onto-genetic origin of the latter, as opposed to the phylogenetic origin of the former. More specifically, learned behavior presupposes a mechanism of learning: the law of effect, the show-piece of behaviorism. This law of effect is itself a functional product of evolution and it produces behavior which is, as a rule, functional.

In human beings (and higher animals ?) all this has culminated in the possibility of conscious functional behavior: intentional behavior, again a functional product of evolution. The capacity of intentional behavior obviously constitutes the kernel of so-called human freedom. As we have seen before, intentional explanation of actions, and hence human freedom, does not compete with nomological explanation at all.

We do not, of course, want to suggest that all non-instinctive human behavior is intentional. On the contrary, the law of effect remains to produce unconscious functional behavior: the so-called second nature, to be distinguished from the first nature of physiological processes and instinctive behavior. But, unfortunately, not all second nature behavior is only functional, especially in

cases where 'unconscious goals and motives' are working which are in conflict with conscious goals. This is the domain of psychological explanation and therapy. Whereas behavioral therapy bets on the law of effect in the attempt to stop or change problematic second-nature behavior, so-called critical therapy bets on making conscious, as the proper way to increase the amount of undisturbed intentional behavior.

It is tempting to extrapolate our general diagnosis of explanatory arguments to the social level. To be sure, in the context of so-called institutional subjects, like firms and states, it generally makes sense to explain their behavior intentionally. But apart from that there is also the phenomenon of collective effects of individual intentional behavior, for example, utilistic choice-behavior. Such effects may or may not be functional on the level of groups or whole societies. However this may be, we arrive again at a type of phenomena, also occurring in non-life sciences and population-biology, where straightforward nomological explanations are feasible and interesting, namely collective results of in some respects and, as a rule, similar behavior of many individual entities of one kind or another. Such explanations will, as a rule, not postulate goals or functions at the group-level. Consequently, they do not imply intentional or functional explanations on this level.

NOTES

(1) See Carl G. Hempel, 'Aspects of scientific explanation', Section 10, in his *Aspects of scientific explanation*, Free Press, 1965.

(2) See Georg H. von Wright, *Explanation and understanding*, Routledge, 1971. See also his 'Determinism and the study of man' in *Essays on explanation and understanding* (eds. Manninen and Tuomela), Reidel 1976.

(3) We will restrict our attention to a naive version of the nomological theory. Hempel introduces the additional premise that x is a rational agent, with the corresponding change in the nomological premise. But this is not important for our explicative evaluation, i.e. all four objections to be presented remain applicable.

(4) Note that these assumptions do not imply that the two specific premises are mutually independent. In this connection Hempel speaks of the 'epistemic interdependence of belief attributions and goal attributions'.

(5) The use of the term 'standard' in the context of the QA-scheme should not be confused with the expressions 'standard theory' and 'standard assumption' defined in the introduction.

(6) Equivalent 'complete' expressions are e.g. '*x* performed *a* in order to approach goal *g*' and '*x* aimed with *a* at goal *g*'.

(7) To be sure Hempel transforms for this problem the nomological argument into a 'utilistic' argument where the specific goal disappears and is replaced by a number of different outcomes. In this way the explanation of an action gets confused with the explanation of the choice of an action.

(8) Although von Wright is not very clear in this respect, we take the following account as the weakest, and the least objectionable, account of what he wants to say. In general, we have much sympathy with his 'negative account', but we consider his 'positive account' as an unfortunate enterprise.

(9) In what follows it should be remembered that we announced our intention to use *BU(x,a,g)* always in the non-trivial sense, hence also if the '*g*' is replaced by a goal-*variable*.

(10) Similar qualifications may be made for *D(x,g)* and *BU(x,a,g)*, now pertaining to the qualification that *x*'s desire and belief were conscious.

(11) After writing the first version of this paper I came to know, thanks to Theo de Boer, that Rex Martin has considered the possibility that the conclusion of an intentional explanation might be that '*x* performed *a* intentionally'. But he rejects it, in our opinion, mainly due to the fact that he does not see that existential generalization is involved. Cf. Rex Martin, 'The problem of the 'tie' in von Wright's schema of practical inference: a Wittgensteinian solution' in *Essays on Wittgenstein in honour of G.H. von Wright*, Acta Philosophica Fennica, Vol. 28, Nos. 1–3. (ed. J. Hintikka), North-Holland, 1976, pp. 326–363, in particular pp. 346–349. According to Martin the idea has also been considered, and rejected, by Norman Malcolm in a (as yet unpublished) paper entitled 'Intention and belief'. It should be conceded, and even stressed, that in many other publications, verbal characterizations of intentional explanation occur which are similar to our analysis. The important point is that the logical structure has not been made explicit in such presentations.

(12) In Section 4 we will hint upon the possibility of 'unconscious goals'.

(13) Here, and elsewhere, the intuitive notion of 'historical explanation' should not be confused with 'intentional explanation in history'.

(14) The following extended interpretation is also plausible. From our article 'Approaching descriptive and theoretical truth' (*Erkenntnis*, 18. 3, 1982, pp. 343–378) it follows that a distinction can be made between descriptive and theoretical sciences. The science of history is obviously a descriptive science; in a sense it is the only one. The nomological criticism is based on the false claim that the science of history is, or should be, a theoretical science, where nomological explanations play an important role. In constrast to this view, intentional explanations can be of crucial importance in the science of history : for it is easy to show that a true intentional explanation, conceived as the transition from *P(x,a)* to *PI(x,a,g)*, is a (synthetic) step which brings us closer to the descriptive truth.

(*) I would like to thank A. van den Beld, Th. de Boer, T. Dehue, D. Draaisma, M. Hoekstra, M. Jeuken, J. Mooij and H. Reddingius for their useful comments on a first (Dutch) version, containing the same idea but differing a lot in presentation. I also thank Illka Niiniluoto for his stimulating critical remarks. I am very grateful to the NIAS (Wassenaar), where my stay as a fellow enabled me to write a rather new version, and to Ann Simpson of the NIAS for improving the English.